

Portland State University PDXScholar

Civil and Environmental Engineering Faculty
Publications and Presentations

Civil and Environmental Engineering

4-1-2012

Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting

Caleb Matthew DeChant
Portland State University

Hamid Moradkhani
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/cengin_fac

 Part of the [Civil and Environmental Engineering Commons](#)

Citation Details

DeChant, C. M., and H. Moradkhani (2012), Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, *Water Resour. Res.*, 48, W04518.

This Article is brought to you for free and open access. It has been accepted for inclusion in Civil and Environmental Engineering Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting

Caleb M. DeChant¹ and Hamid Moradkhani¹

Received 7 June 2011; revised 27 February 2012; accepted 1 March 2012; published 18 April 2012.

[1] In hydrologic modeling, state-parameter estimation using data assimilation techniques is increasing in popularity. Several studies, using both the ensemble Kalman filter (EnKF) and the particle filter (PF) to estimate both model states and parameters have been published in recent years. Though there is increasing interest and a growing literature in this area, relatively little research has been presented to examine the effectiveness and robustness of these methods to estimate uncertainty. This study suggests that state-parameter estimation studies need to provide a more rigorous testing of these techniques than has previously been presented. With this in mind, this paper presents a study with multiple calibration replicates and a range of performance measures to test the ability of each technique to calibrate two separate hydrologic models. The results show that the EnKF is consistently overconfident in predicting streamflow, which relates to the assumption of a Gaussian error structure. In addition, the EnKF and PF were found to perform similarly in terms of tracking the observations with an expected value, but the potential for filter divergence in the EnKF is highlighted.

Citation: DeChant, C. M., and H. Moradkhani (2012), Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, *Water Resour. Res.*, 48, W04518, doi:10.1029/2011WR011011.

1. Introduction

[2] Within the hydrologic modeling community, there are many new and developing perspectives on the methods through which uncertainty should be estimated. The newest techniques were developed due to a recent shift in focus of model calibration from simple optimization to probabilistic characterization of model parameters [Beven and Freer, 2001]. With the recognition of multiple different uncertainty sources (i.e., forcing data, observation, model structure, and parameters), much of the community has tried to account for these uncertainties at varying levels [e.g., Bulygina and Gupta, 2009; Kavetzki et al., 2006; Moradkhani et al., 2006; Moradkhani and Meskele, 2009; Vrugt et al., 2008]. This has led to an array of different probabilistic techniques to estimate the uncertainty in a given modeling framework. Through an analysis of the uncertainty in a model prediction, the ultimate goal is to produce an accurate probabilistic forecast of a given hydrologic variable. An accurate probabilistic forecast is necessary to allow for effective decision making in the management of water resources. Some examples of attempts in the literature to analyze the uncertainty in hydrologic prediction include the generalized likelihood uncertainty estimator (GLUE) [Beven and Freer, 2001; Stedinger et al., 2008], Markov chain Monte Carlo (MCMC) [Jeremiah et al., 2011; Smith

and Marshall, 2008; Vrugt et al., 2008], Bayesian total error analysis [Kavetski et al., 2002], data assimilation [DeChant and Moradkhani, 2011b; Liu and Gupta, 2007; Moradkhani et al., 2005a, 2005b; Moradkhani, 2008], combined data assimilation and Bayesian model averaging [Parrish et al., 2012], and hierarchical Bayesian [Wu et al., 2010] methods. With this collection of methods at hand, there is great potential for improving the handling of uncertainty in hydrologic modeling and improving the accuracy of probabilistic forecasts.

[3] The study presented here focuses on the use of data assimilation techniques to manage the uncertainty in the modeling framework. Of the above-mentioned methods, data assimilation is attractive for a number of reasons. First, the data assimilation framework provides a methodology for handling all sources of modeling error simultaneously. Second, data assimilation is performed sequentially and therefore has potential in an operational framework, where the estimation of hydrologic quantities is desired at regular intervals. The last benefit of data assimilation is that it does not rely on the assumption of stationarity. Through a sequential estimation of parameters, data assimilation has the potential to handle changes in hydrologic flow patterns.

[4] In the hydrologic data assimilation literature, recent studies have examined the estimation of uncertainty in parameters of a hydrologic model, in addition to the more traditional state estimation [Moradkhani and Sorooshian, 2008]. Through the inclusion of parameters in the data assimilation process, it is hypothesized that the total uncertainty in the prediction can be more accurately characterized. Several recent studies of state-parameter estimation in hydrologic models have utilized the popular EnKF [DeChant

¹Department of Civil and Environmental Engineering, Portland State University, Portland, Oregon, USA.

and Moradkhani, 2011a; Franssen and Kinzelbeck, 2008; Leisenring and Moradkhani, 2011; Moradkhani et al., 2005b; Wang et al., 2009]. In addition to the EnKF, particle filters (PF) have been increasing in popularity for both state and state-parameter estimation [DeChant and Moradkhani, 2011a; Leisenring and Moradkhani, 2011; Montzka et al., 2010; Moradkhani et al., 2005a; Nagarajan et al., 2010; Rings et al., 2010; Salamon and Feyen, 2009; Smith et al., 2008; Weerts and El Serafy, 2006]. Of the recent attention that has been paid to state-parameter estimation in the EnKF and PF, little has been shown as to the robustness of these two techniques. It is necessary for the hydrologic data assimilation community to address the effectiveness of both techniques for state-parameter estimation over different scenarios to prove the applicability of the techniques, and relate the results back to the statistical theory and their inherent assumptions. This study aims to perform such an analysis with two conceptual rainfall-runoff models of differing complexities. Throughout this analysis, the importance of examining the behavior of techniques over many different scenarios is highlighted. This study is organized as follows. Section 2 discusses the formulation of the data assimilation techniques and the study basin. Section 3 discusses the experimental setup, including the hydrologic models, time-lagged replicates of the experiment, and the methods through which these replicates are validated. Section 4 presents the results of the data assimilation techniques followed by a discussion of the results and the conclusion in section 5.

2. Data Assimilation Techniques

2.1. Ensemble Kalman Filter

[5] The EnKF is an ensemble version of the Kalman filter, performed as a Monte Carlo simulation, in order to overcome the need for a linear model (Kalman filter) and the need to obtain the derivative of the model for calculation of the error covariances (extended Kalman filter) [Evensen, 2003]. Through an ensemble framework, the need for model linearization is relaxed and the error covariances can be calculated from the ensembles [Moradkhani et al., 2005b]. Implementation of the EnKF begins at the initial time step of modeling. At this initial time step, the model is supplied with an initial distribution of states and parameters. As the model progresses forward in time, the prior distribution of states is produced according to equation (1):

$$x_{i,t}^- = f(x_{i,t-1}^+, u_{i,t}, \theta_{i,t}^-) + \omega_{i,t}, \quad (1)$$

where f is the forward operator (hydrologic model), $x_{i,t}^-$ represents the model predicted (prior) states, $x_{i,t-1}^+$ represents the posterior model states at the previous time step, $u_{i,t}$ represents the meteorological forcing data, $\theta_{i,t}^-$ represents the prior model parameters at the current time step, $\omega_{i,t}$ represents the model error, i is the ensemble member, and t is the time step. In order to describe parameter estimation, it is also necessary to describe the estimation of the prior parameter distribution, which is shown in equation (2):

$$\theta_{i,t}^- = \theta_{i,t-1}^+ + \eta S(\theta_{\bullet,t-1}^-), \quad (2)$$

where η is a hyper-parameter to retain diversity in parameters, which was tuned to 0.01 for this application, and

$S(\theta_{\bullet,t-1}^-)$ is the standard deviation of the prior parameter distribution at the previous time step ($\theta_{\bullet,t-1}^-$). In addition, the prior parameter distribution at the initial time step is developed using Latin-hypercube sampling. Prior to update of the model states and parameters, an observational operator must be applied to transfer the states into the observation space, as in equation (3):

$$y_{i,t}' = h(x_{i,t}^-, \theta_{i,t}^-) + \nu_{i,t}, \quad (3)$$

where $h(\bullet)$ is the observational operator (hydrologic routing), which translates the surface water and storages to flow at the watershed outlet, $y_{i,t}'$ is the prediction (stream-flow), and $\nu_{i,t}$ represents the prediction error. After the prediction is obtained, the posterior states and parameters are estimated with the Kalman update equation as follows [Moradkhani, 2008]:

$$x_{i,t}^+ = x_{i,t}^- + K_{xy}[y_t + \varepsilon_{i,t} - y_{i,t}'], \quad (4)$$

$$\theta_{i,t}^+ = \theta_{i,t}^- + K_{\theta y}[y_t + \varepsilon_{i,t} - y_{i,t}'], \quad (5)$$

where y_t is the observed flow, $\varepsilon_{i,t}$ represents the observation error, and K_{xy} and $K_{\theta y}$ are the Kalman gains for states and parameters, respectively. The Kalman gain in state space is calculated from equation (6):

$$K_{xy} = P^- H^T (H P^- H^T + R_t)^{-1} = C_{xy} (C_{yy} + R_t)^{-1}. \quad (6)$$

In equation (6), $P^- H^T = C_{xy}$ is the covariance of the states ensemble with the predicted observation, $H P^- H^T = C_{yy}$ is the variance of the predicted observations, and R_t is the observation error variance at time t . H is the linearized observation operator ($H = \frac{\partial h}{\partial x}$). The Kalman gain for the parameters can be obtained similar to equation (6), as shown by Moradkhani et al. [2005b]. The model state error covariance P^- can now be computed directly from the ensemble deviations (e^-):

$$P^- = \frac{1}{N_{\text{ens}}} e^- e^{-T}, \quad (7)$$

$$e_i^- = x_{i,t}^- - \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} x_{i,t}^-, \quad (8)$$

$$K_{xy'} = C_{xy'} (C_{yy} + R_t)^{-1}, \quad (9)$$

where N_{ens} is the ensemble size.

2.2. Particle Filter

[6] The PF, similar to the EnKF, sequentially calculates the posterior distribution of states and parameters. The advantage of the PF, in comparison to the EnKF, is that it relaxes the assumption of a Gaussian error structure, which allows the PF to more accurately predict the posterior distribution in the presence of skewed distributions [Moradkhani et al., 2005a]. This is accomplished by resampling sets of state and parameters, or “particles,” with higher posterior weights, as opposed to the linear model state updating of

the EnKF. The PF used in this study is the sequential importance resampling (SIR) PF. Since SIR is used in this study, the PF will be referred to as PF-SIR to be specific to the method while presenting the results.

[7] Based on the recursive Bayes Law (equation (10)), the PF sequentially samples prior states and parameters to create an accurate posterior distribution, at each observation time step,

$$p(x_t, \theta_t | y_{1:t}) = p(x_t, \theta_t | y_t, y_{1:t-1}) = \frac{p(y_t | x_t, \theta_t) p(x_t, \theta_t | y_{1:t-1})}{\int p(y_t | x_t, \theta_t) p(x_t, \theta_t | y_{1:t-1}) dx_t} \quad (10)$$

Equation (10) shows mathematically that the posterior distribution of model-predicted states (x_t) and parameters (θ_t), given the observations ($y_{1:t}$), can be computed sequentially in time. In this study, the probability of each particle is calculated via the normal likelihood equation (11):

$$L(y_t | x_{i,t}^-, \theta_{i,t}^-) = \frac{1}{\sqrt{2\pi R_t}} \exp\left(-\frac{1}{2R_t} [y_t - y'_{i,t}]^2\right). \quad (11)$$

The normalized likelihood, $p(y_t | x_{i,t}^-, \theta_{i,t}^-)$, can easily be calculated by:

$$p(y_t | x_{i,t}^-, \theta_{i,t}^-) = \frac{L(y_t | x_{i,t}^-, \theta_{i,t}^-)}{\sum_{i=1}^{N_{\text{ens}}} L(y_t | x_{i,t}^-, \theta_{i,t}^-)} = p(y_t - y'_{i,t} | R_t). \quad (12)$$

This probability is necessary to transform the prior particle weights into the posterior via equation (13):

$$w_{i,t}^+ = \frac{w_{i,t}^- \cdot p(y_t | x_{i,t}^-, \theta_{i,t}^-)}{\sum_{i=1}^{N_{\text{ens}}} w_{i,t}^- \cdot p(y_t | x_{i,t}^-, \theta_{i,t}^-)} \quad (13)$$

In the PF-SIR, prior particle weights, $w_{i,t}^-$, are set equal to $1/N_{\text{ens}}$ before moving on to the next time step. This results in a posterior weight, $w_{i,t}^+$, equal to $p(y_t | x_{i,t}^-, \theta_{i,t}^-)$, which is the normalized likelihood. The SIR algorithm resamples the states with a probability greater than uniform probability. *Leisenring and Moradkhani* [2011] examined weighted random resampling (WRR) in comparison with SIR for the SNOW-17 model and concluded marginal improvement in the performance of the PF. In this study, the SIR method is implemented as elaborated by *Moradkhani et al.* [2005a].

3. Experimental Setup

3.1. Case Study: Leaf River Basin

[8] This study takes place over the Leaf River Basin in southern Mississippi. The basin is 1944 km² and is the main tributary of the Pascagoula River, which drains into the Gulf of Mexico. Data for this study was obtained from the National Weather Service Hydrology Laboratory, which consists of precipitation (mm d⁻¹), potential evapotranspiration (mm d⁻¹), and streamflow (cm³ s⁻¹). This data set has observations from October 1948 through September 1988

providing 40 yr of data for analysis. The methods for utilizing the entire data set are described in section 3.2. A map of the Leaf River Basin is presented in Figure 1.

3.2. HyMod Model

[9] The HyMod model is a simple, conceptual, lumped model containing five calibration parameters. Based on these five parameters, the model allocates water between a series of three quick-flow tanks and one slow-flow tank, then routes the runoff to the outlet. A description of the parameters, and the possible range of their values, is provided in Table 1. In addition to the parameters, all five HyMod states are estimated as well. For a more detailed description of the model processes, see *Moradkhani et al.* [2005a].

3.3. Sacramento Soil Moisture Accounting (SAC-SMA) Model

[10] The SAC-SMA model, first introduced by *Burnash et al.* [1973], is a conceptual water balance model used operationally at the National Weather Service River Forecast Center. The model simulates water storage with two soil moisture zones: an upper and a lower zone. The upper zone accounts for short-term storage of water in the soil, while the lower zone models the longer-term groundwater storage. Water can move vertically from the upper zone to the lower zone, laterally out of the system depending on the state variables and the parameterization, or vertically out of the system through evapotranspiration. Excess runoff is routed to the watershed outlet using a Nash cascade of three linear reservoirs. The SAC-SMA model parameters are summarized in Table 2. In addition to these parameters, all six SAC-SMA states and the storage in the three Nash-cascade reservoirs are estimated.

3.4. Assumed Errors

[11] In any data assimilation framework, it is necessary to assume error values for any quantity that contains uncertainties. This study applies noise directly to the precipitation, potential evapotranspiration (PET), model predictions, and streamflow observations to account for their uncertainties. Precipitation is assumed to have a lognormal error distribution with a relative error of 25%. Similarly, PET error is assumed to follow a normal distribution with a relative error of 25%. Both these values are necessary to account for errors in meteorological measurements due to spatial heterogeneity of these variables and sensor errors. All prediction errors are assumed to be normally distributed with a relative error of 30% for HyMod and 25% for SAC-SMA. Differing values for these models reflects the different accuracies in streamflow prediction. Last, the streamflow observation errors are assumed to be normally distributed with a relative error of 15%. All errors in this study are assumed to be uncorrelated. Errors assumptions are applied with the same magnitude in both the EnKF and PF-SIR. The assumed values were determined through a manual tuning to achieve the most reliable predictions over the time-lagged calibration replicates. Though the assumed values were calibrated to achieve the most reliable ensemble prediction over the entire observation period, it is necessary to caution the reader that these are not necessarily the physically correct error terms. Since these errors were determined with very little a priori knowledge about the real



Figure 1. The Leaf River Basin in southern Mississippi.

error magnitudes, their estimation is ill-posed, as explained by *Renard et al.* [2010], and therefore are uncertain.

3.5. Time-Lagged Calibration Replicates

[12] In order to examine the robustness of both the EnKF and PF, this study implements each assimilation technique in time-lagged calibration periods between October 1948 and August 1981. The setup of the time-lagged calibration periods is shown in Figure 2. Figure 2 shows the 21 different calibration time periods (lagged by 500 d) for the HyMod model and the 10 different calibration time periods (lagged by 1000 d) for the SAC-SMA model. Each model calibration is 2000 time steps, and assimilates streamflow observations at a daily frequency. During each separate calibration time period, performance measures were only calculated for the second half (1000 d) of the model run to allow for states and parameters to converge to plausible values. The smaller number of calibration replicates for the SAC-SMA model were used because of increased computational time and greater required storage space due to the increased number of states and parameters. Following the model calibrations, the posterior parameters at the last time step of each calibration are used to run the model during a validation period from September 1981 through

January 1987. Each validation is performed with state-only estimation, using the estimated posterior parameter distribution from the calibration. During these validation experiments, all noise terms are consistent with the calibration except that no parameter perturbation or evolution is considered. This is because the parameter distribution is assumed to be constant during the validation. Performing the validation is intended to assess the performance of the posterior parameter distribution, created by each method, on an independent data set. This provides insight into the accuracy of the posterior parameter distribution.

3.6. Deterministic and Probabilistic Performance Assessment

[13] In order to provide a robust analysis of each assimilation run, it was necessary to calculate multiple performance measures. Four quantitative measures and two graphical measures were used to check assimilation performance. The first is the Nash-Sutcliffe efficiency (NSE), which is the only measure of the accuracy of the expected value (EV). This shows the ability of each technique to track the observation. In terms of probabilistic measures, the normalized root-mean-square error ratio (NRR), 95% exceedance ratio (ER95) [Moradkhani et al., 2006], reliability (α), rank histogram, and quantile-quantile plot (Q-Q plot) were examined. All probabilistic measures are ensemble verification techniques over a time series of observation (i.e., streamflow). It is important to note that α is a measure of the proximity of the Q-Q plot to uniform, which was suggested by *Renard et al.* [2010]. *Renard et al.* [2010] also proposed a second reliability score, (ξ), which measures the percentage of observations falling within the ensemble prediction. In the analysis of this experiment, ξ is not utilized because the ER95 provides similar information. All measures are described in Table 3. Each of

Table 1. HyMod Model Parameter Descriptions With Feasible Ranges

Parameter	Description	Range
R_q	Quick flow tank parameter	0–1
R_s	Slow flow tank parameter	0.001–0.1
Alpha	Partitioning factor	0.6–1
Beta	Variability of soil moisture capacity	0–2
C_{\max}	Maximum watershed storage capacity	0–1000

Table 2. SAC-SMA Model Parameters Description With Feasible Ranges

Parameter	Description	Units	Range
Capacity parameters			
UZTWM	Upper zone tension water maximum	(mm)	10–300
UZFWM	Upper zone free water maximum	(mm)	5–150
LZTWM	Lower zone tension water maximum	(mm)	10–500
LZFPM	Lower zone free primary maximum	(mm)	10–1000
LZFSM	Lower zone free secondary maximum	(mm)	5–400
ADIMP	Additional impervious area	–	0–0.4
Recession parameters			
UZK	Upper zone depletion parameter	1 d ⁻¹	0.1–0.75
LZPK	Lower zone primary depletion parameter	1 d ⁻¹	0.0001–0.05
LZSK	Lower zone secondary depletion parameter	1 d ⁻¹	0.01–0.35
Percolation and other			
ZPERC	Maximum percolation rate	–	5–350
REXP	Percolation equation exponent	–	1–5
PCTIM	Impervious area of watershed	–	0–0.1
PFREE	Free water percolation from upper to lower zone	–	0–0.1
Routing parameter			
Kq	Nash-cascade routing parameter	1 d ⁻¹	0.1–0.5
Not estimated			
RIVA	Riparian vegetated area	–	
SIDE	Deep recharge to channel base flow	–	
RSERV	Lower zone free water not transferable to tension water	–	

these performance metrics examines the ability of prior streamflow forecasts to predict the observed streamflow.

4. Results

4.1. HyMod Results

[14] Each calibration of the HyMod model was performed with different ensemble sizes from 10 to 1000. By applying the EnKF and PF-SIR with 15 different ensemble sizes, the performance of each assimilation technique, with respect to ensemble size, is analyzed. In order to display the results of all model calibrations, the performance measures for all 21 different lagged calibration periods are averaged at each ensemble size and plotted in Figure 3.

[15] Over the four subplots in Figure 3, some contradictory results are observed. First, it is noted that the EnKF produces a greater NSE and α than the PF-SIR at all ensemble sizes. This greater NSE and α suggest that the EnKF produced a more accurate expected value and a more reliable ensemble prediction than the PF-SIR. Though this suggests the EnKF is more effective than the PF-SIR, the NRR and ER95, suggest different results. The NRR indicates a perfect characterization of uncertainty when equal to one, while values less than one indicate too much spread (underconfident) in the ensemble prediction and values greater than one indicate an ensemble with too little spread (overconfident). Also, the ER95 will be 5% for an ideal distribution. ER95 greater than five suggests the distribution is too narrow and ER95 less than five suggests the distribution is too wide. With an NRR closer to one and ER95 closer to 5%, the PF-SIR appears to have produced a more accurate characterization of the uncertainty. The EnKF produced a more accurate prediction but had a stronger tendency to be overconfident (the uncertainty in the system is routinely underestimated). This suggests that, although the EnKF predicted the mode of the posterior distribution more accurately, it struggled to estimate the tails of the posterior distribution. In comparison, the PF-SIR was less accurate in

estimating the mode of the posterior distribution but more accurate in estimating the tails. In order to verify that the averaged results of Figure 3 are representative of all 21 calibration time periods, Figure 4 is presented to show the performance measures of each time-lagged model calibration.

[16] Figure 4 provides validation that the averaged results of Figure 3 are representative of all 21 time-lagged model calibrations. The NSE and α subplots show that the EnKF produced a more accurate expected value and reliable ensemble prediction in nearly every model calibration, confirming the results found in Figure 3. Although the mode of the posterior distribution is more accurately characterized by the EnKF than the PF-SIR, the NRR and ER95 subplots confirm that the EnKF had a greater tendency toward overconfidence than the PF-SIR. This also confirms the conclusion that the PF-SIR more accurately characterized the tails of the posterior distribution. This is an important observation with respect to the assessment of uncertainty in hydrologic forecasting. In order to accurately estimate the uncertainty in the modeling framework, it is necessary to accurately estimate the entire posterior distribution. The need for accurate estimation throughout the posterior is discussed further in the analysis of subsequent figures.

[17] While the performance measures presented thus far provide information into the general accuracy of the predictive distribution, a visual representation is necessary to provide further insight into the behavior of each model. Therefore, we further this analysis by visualizing the data in the form of rank histograms and Q-Q plots in Figure 5. The rank histograms and Q-Q plots presented here show the results from the predictive distributions with an ensemble size of 1000 for the last 1000 d of each model calibration, for both the EnKF and PF-SIR. By examining the rank histograms, it is observed that both the EnKF and PF-SIR have a large quantity of observations that fall in the outer bins of the distribution, indicating an overconfidence problem for each method. While this confirms the previous results, it also provides information on the bias in the

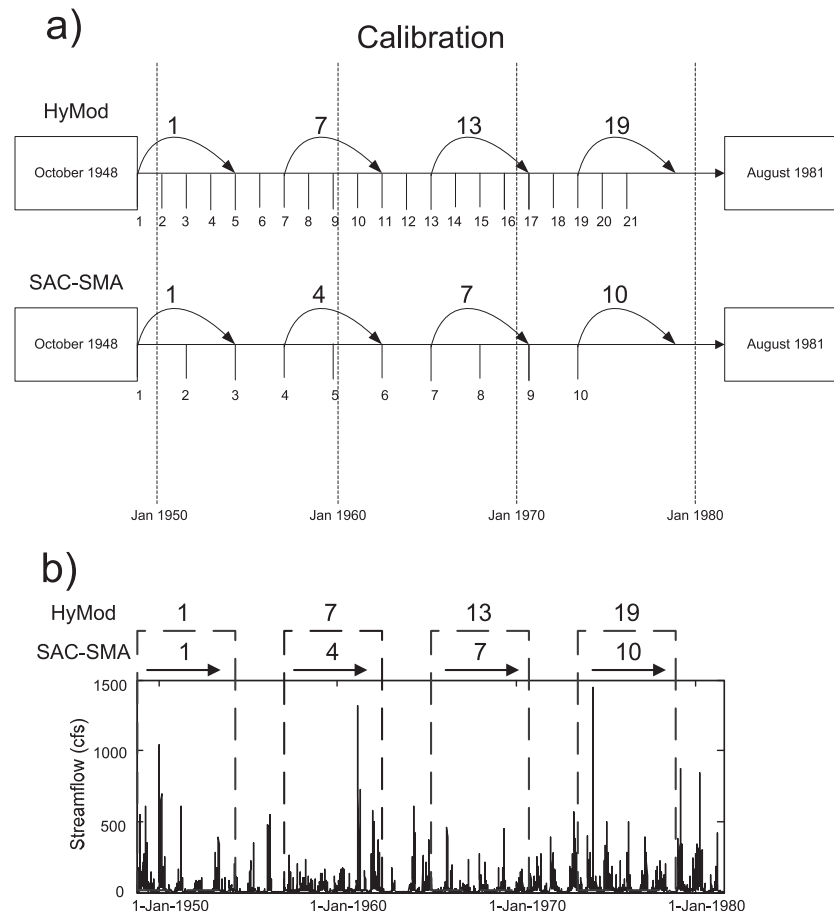


Figure 2. Schematic of the lagged calibration time periods for the joint state-parameter estimation of the HyMod and SAC-SMA models. (a) The start time of each calibration. The curved arrows show the model run time and correspond to the numbered boxes in subplot b. (b) The hydrograph over the entire calibration period with boxes showing four sample calibration periods (1, 7, 13, 19 for HyMod and 1, 4, 7, 10 for SAC-SMA).

predictive distribution, which was not previously addressed. Each method tended to overpredict the low flows, indicating the model struggles to predict low flows. Though both methods struggled with the low flows, only the EnKF underpredicted the high flows. This poor characterization of the high flows caused the higher NRR and ER95 values observed in Figures 3 and 4. In addition to the rank histograms, the Q-Q plots also indicate a tendency toward overprediction in the low flows using both techniques, and an underprediction of the high flows in the EnKF. The Q-Q plot shows how bias can make the interpretation of α difficult. Since the Q-Q plot for the EnKF crosses the uniform line, it actually produces a higher α , but provides a worse ensemble prediction than the PF-SIR. As was suggested previously, it is observed that the EnKF struggled to predict the posterior tails, particularly the tail producing the high flows, in comparison to the PF-SIR. In general, this poor estimation of the full posterior by the EnKF is caused by the assumption of a Gaussian error structure. In predicting streamflow, highly skewed error structures are quite common, especially in models as simple as HyMod. In the presence of non-Gaussian error structures, the EnKF still has the potential to predict the mode of the distribution, but

is incapable of estimating the full posterior distribution. As has been observed thus far, the EnKF accurately predicted the mode of the posterior, but struggled in comparison to the PF-SIR in characterizing the full uncertainty, suggesting the PF-SIR is a more robust uncertainty estimator. This is the result that should be expected of a simple rainfall-runoff model.

[18] Up to this point, the ability of the EnKF and PF-SIR to estimate streamflow during calibration has been assessed, but the ability of each technique to estimate a posterior parameter distribution must be analyzed separately. In order to determine the accuracy of the posterior parameters from each data assimilation technique, a validation of each model calibration is presented. Similar to Figure 3, validation performance measures, averaged over 15 different ensemble sizes, are shown in Figure 6. Note the large differences between the calibration results in Figure 3 and validation results in Figure 6. Overall, from Figure 6 it is observed that all four performance measures indicate the PF-SIR was more accurate for ensemble sizes above 200. While the EnKF more accurately predicted the mode of the predictive distribution during calibration, the results from the validation suggest the PF-SIR performed better in all measures. In

Table 3. Summary of Performance Measures

Performance Measure	Mathematical Representation	Description
Nash-Sutcliffe efficiency (NSE)	$1 - \frac{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y}_{\bullet,t})^2}{\text{var}(y_{1:T})}$	A NSE equal to 1 is a perfect prediction, while a value of 0 indicates no skill beyond the streamflow variability.
Reliability (α)	$1 - \frac{2}{T} \sum_{t=1}^T \text{abs}\left(\frac{z_t}{T} - U_t\right)$ See Q-Q plot for description of z_t calculation	A measure of the fit of the quantile plot to uniform. A value of 1 is exactly uniform and a value of 0 is the furthest possible from uniform.
Normalized root-mean-square error ratio (NRR)	$\frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y}_{\bullet,t})^2}}{\frac{1}{N} \left\{ \sum_{i=1}^N \sqrt{\frac{1}{T} \left[\sum_{t=1}^T (y_t - \bar{y}_{\bullet,t})^2 \right]} \right\} \sqrt{\frac{N+1}{2N}}}$	A measure of the spread of the ensemble in relation to the accuracy of the EV. A value of 1 is accurate spread, >1 is a narrow distribution, and <1 is a wide distribution.
95% exceedance ratio (ER95)	$\frac{1}{T} \sum_{t=1}^T (y'_{97.5\%,t} < y_t \text{ or } y'_{2.5\%,t} > y_t) * 100\%$	A perfect ensemble would have a 5% exceedance of the 95% predictive bounds.
Rank histogram	Rank all observations by their location in the sorted (ascending) ensemble $\text{Rank}_t = I(y'_{i,t} < y_t \text{ or } y'_{i+1,t} > y_t)$ A histogram is created of all time steps.	A uniform histogram indicates accurate representation of uncertainty. For a detailed description of rank histogram interpretation, see <i>Hamill</i> [2001].
Quantile-Quantile plot (Q-Q plot)	Calculate the quantile of every observation time step $z_t = \frac{I(y'_{i,t} < y_t \text{ or } y'_{i+1,t} > y_t)}{N}$ Sort the Q-Q matrix and compare with uniform distribution	A Q-Q plot matching the uniform line indicates optimal ensemble prediction. For details of the interpretation of a Q-Q plot, see <i>Laio and Tamea</i> [2007]

order to understand the cause of this shift in results from calibration to validation, the reader is reminded that a cross validation is performed, in which parameters from each calibration are applied to an independent data set. Since the parameters are applied to an independent data set, with a slightly different flow regime, it is essential to accurately

estimate the parameter uncertainty. From the previous results, it is understood that the EnKF had a stronger tendency toward overconfidence than the PF-SIR. This overconfidence has led to overfitting, or underestimation of the uncertainty with respect to the posterior parameters from each time-lagged calibration. While the overfitting of these

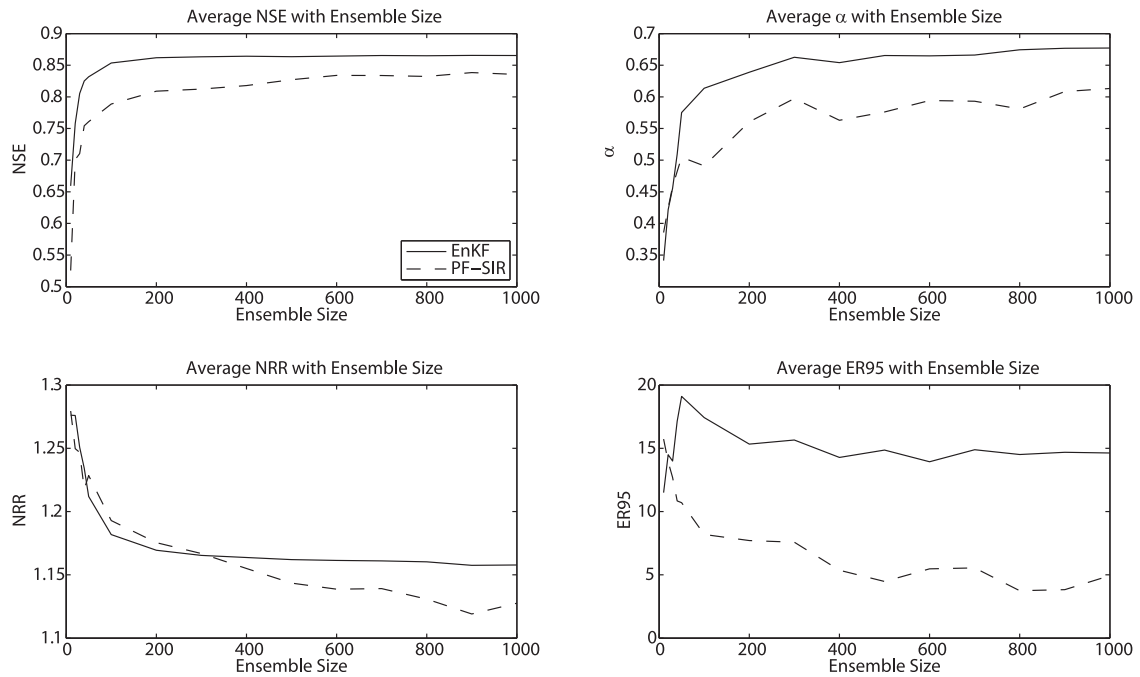


Figure 3. Average verification statistics over all 21 time-lagged calibration periods of the HyMod model with the EnKF (solid lines) and PF-SIR (dotted lines) as a function of the ensemble size.

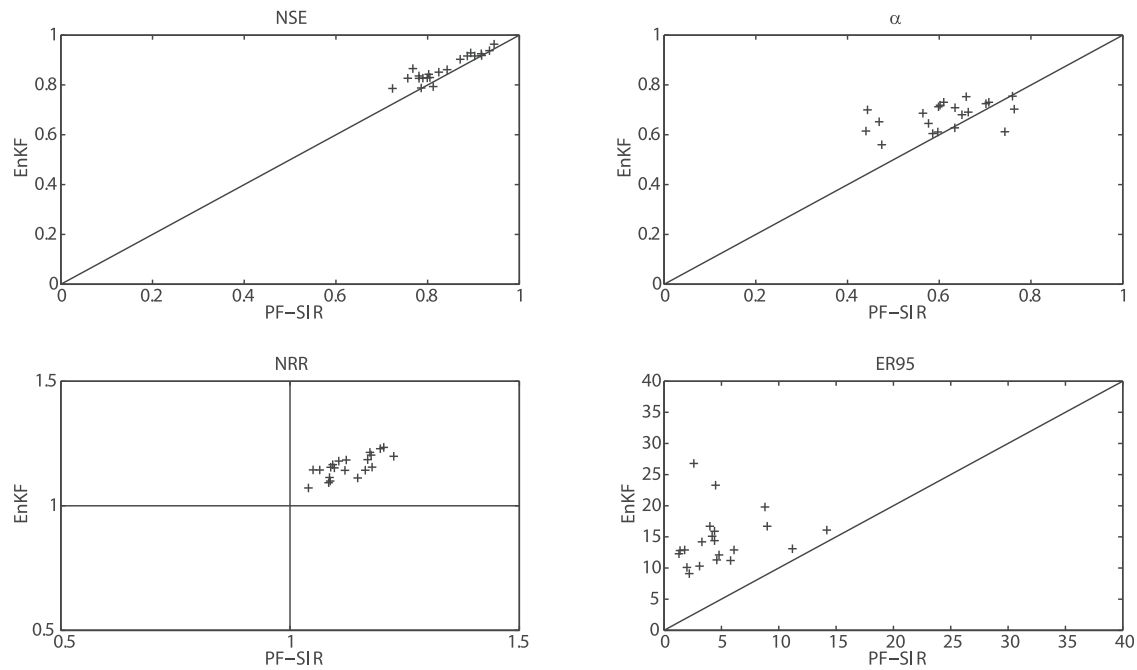


Figure 4. Scatterplot of the performance measures for each of the 21 different time-lagged calibration periods for the HyMod model with the PF-SIR on the x axis and EnKF on the y axis

parameters seemed to be beneficial in streamflow prediction accuracy during the calibration, the validation highlights the negative effects of an incomplete characterization of parameter uncertainty.

[19] To illustrate the consistency of the previous results, scatterplots of the performance measures for the validation are presented in Figure 7. Results in this figure show that the PF-SIR was not only more accurate than the EnKF in terms of expected value and ensemble prediction, for nearly

every validation the total uncertainty was more accurately estimated.

[20] Further insight into the behavior of these two techniques is observed through rank histograms and Q-Q plots in Figure 8. First, and most importantly, differences between the calibration and validation rank histograms must be evaluated. During the validation time steps, the overconfidence in the EnKF prediction appears to be exacerbated. Both the lower and upper bins of the histogram are taller

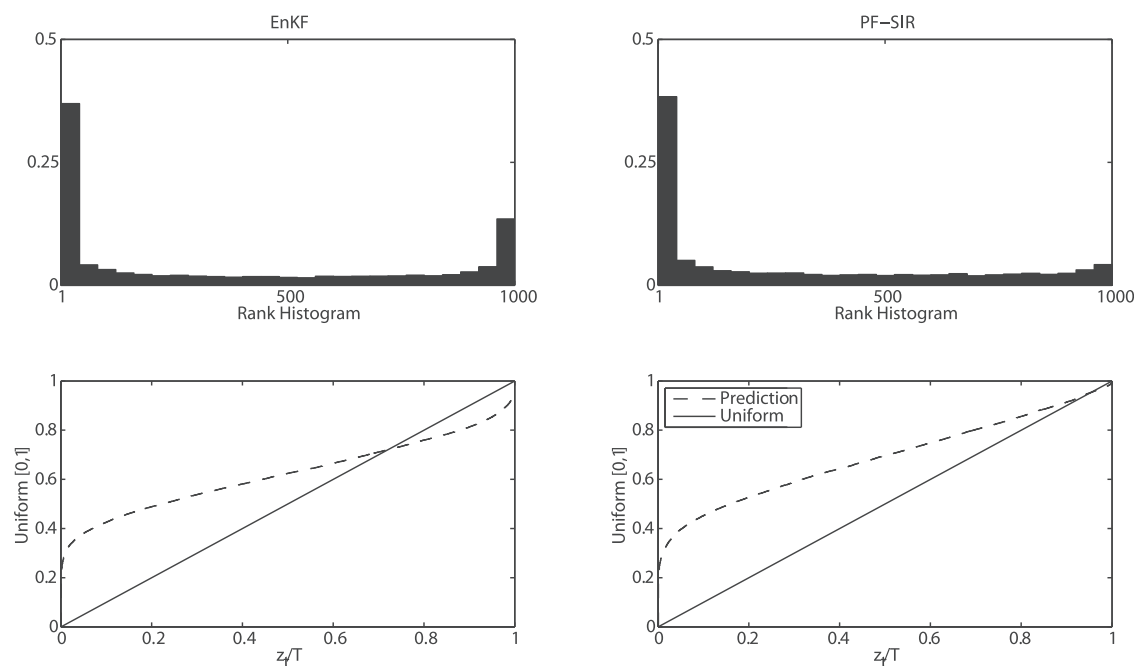


Figure 5. Rank Histogram (top) and Q-Q plot (bottom) for the EnKF (left) and PF-SIR (right) for the HyMod calibration runs with 1000 ensemble members.

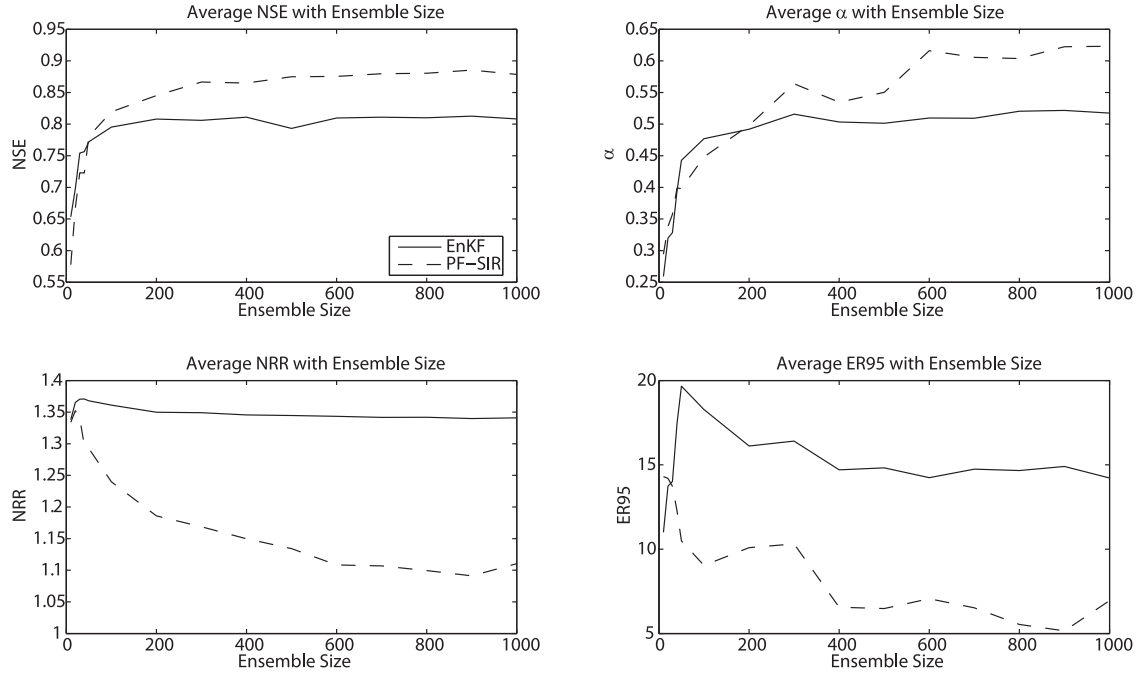


Figure 6. Average performance measures over all 21 validation runs of the HyMod model with the EnKF (solid line) and PF-SIR (dashed line) as a function of the ensemble size.

during validation than the calibration. Further, over half of the observations fell below the predictive distribution during validation. In addition, a highly overconfident prediction is indicated by the flat Q-Q plot. While the EnKF became more overconfident during the validation, the PF-SIR estimated the uncertainty with similar accuracy to the calibration. This again suggests that the posterior parameters produced by the PF-SIR are a more accurate representation of the uncertainty than the posterior created by the EnKF. This is examined further in Figure 9, which shows the combined posterior distribution of each parameter from all calibration replicates. From Figure 9, it is clear that the EnKF converges to a smaller posterior distribution than the PF-SIR for each parameter. In conjunction with the streamflow results, this provides evidence that the EnKF poorly estimated the full posterior parameter distribution. Overall, the EnKF appears to be overfitting the parameters to the data during the calibration runs. A poor characterization of the full posterior in the EnKF is a result of the skewed error structure during the lagged calibrations. Since the EnKF assumes a normal error structure, the tails of the posterior distribution are incorrectly estimated, leading to narrow parameter distributions. This skewed error structure is theoretically less problematic in the PF-SIR, and the results support that.

4.2. SAC-SMA Results

[21] In section 4.2, results from the SAC-SMA model are presented. The additional model analysis is necessary for two reasons. First, it is important to analyze the effects of greater model complexity on the performance of each method. Second, using a different model allows for analysis of each technique's behavior under a different model structure. Similar to the HyMod model, the SAC-SMA was calibrated over each time-lagged period and analyzed with NSE,

α , NRR, and ER95. In the SAC-SMA analysis, results of the performance, with respect to ensemble size, were similar to the HyMod results. In the interest of simplifying the results presentation, the quantitative performance measures are summarized in Table 4 for the 2000 ensemble member case.

[22] The results in Table 4 provide a few contradictory results in comparison to those found during the HyMod model calibration. First, the NSE suggests that the PF-SIR more accurately reproduced the observation than the EnKF during the calibration runs. This result is skewed by a single poor calibration by the EnKF (December 1967–May 1973), but in most calibration runs the expected value from the EnKF was nearly equivalent to the PF-SIR. During the single poor calibration run, filter divergence occurred in the EnKF. This case of filter divergence hints that the EnKF may be less robust than the PF-SIR, in terms of parameter estimation, and this is discussed in detail in section 4.3. In comparison to the HyMod model, the accuracy of the ensemble prediction is also different for the calibration. Unlike the NSE, the average α value during the calibration replicates is not subject to an outlier. This highlights the importance of model structure in the comparison of the EnKF and PF-SIR. Model structure strongly affects the ability of data assimilation techniques to update model states and parameters. This point is discussed further in section 5.1. In terms of ensemble spread and 95% predictive bounds, consistent results are found in comparison to the HyMod model. During the calibration, results suggest that both techniques have a tendency toward overconfidence, but to a lesser extent than in the HyMod results. A consistent trend is observed of the EnKF producing results that are more overconfident than the PF-SIR.

[23] In order to further examine the uncertainty estimation in the SAC-SMA model, the rank histograms and Q-Q plots are provided as well. From Figure 10, it is important

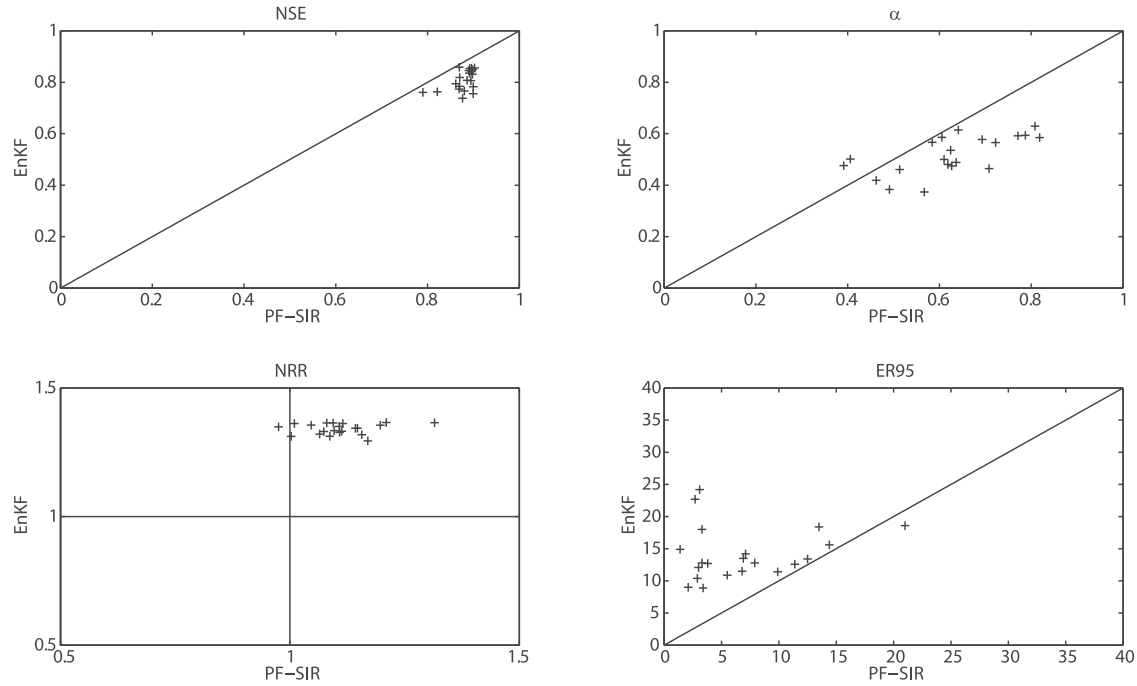


Figure 7. Scatterplot of the performance measures for each of the 21 validation runs for the HyMod model with the PF-SIR on the x axis and EnKF on the y axis.

to note that the accuracy of the SAC-SMA during extreme flow events is quite different from the HyMod model. In the HyMod model, both methods showed difficulty predicting the low flows, but in the SAC-SMA model both techniques have difficulty predicting the high flows. Though both techniques struggled to predict the high flows, the problem is amplified when using the EnKF. Overconfidence persists when using the EnKF to calibrate the SAC-SMA model,

indicating that the error structure in the SAC-SMA model predictions is sufficiently skewed to violate the Gaussian assumption. Though the assumption is violated, the results appear to be less adverse than in the HyMod model, which is likely a result of generally more accurate predictions from the SAC-SMA model.

[24] To analyze the performance of the EnKF and PF-SIR in estimating the posterior parameters in the SAC-SMA

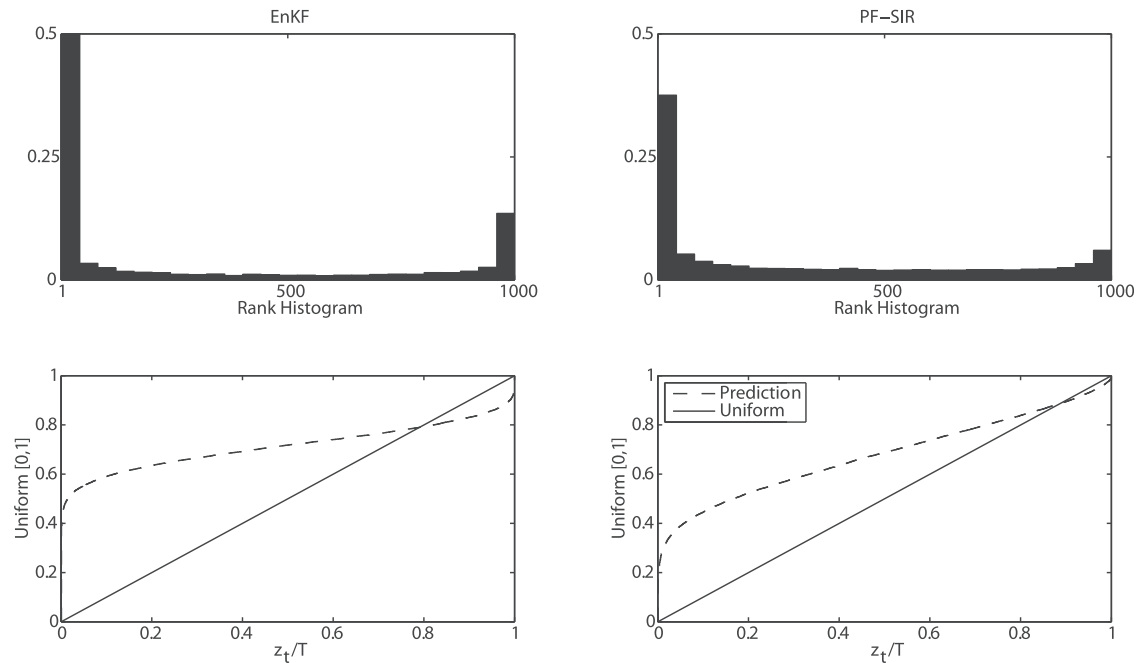


Figure 8. Rank histogram (top) and Q-Q plot (bottom) for the EnKF (left) and PF-SIR (right) for the HyMod validation runs with 1000 ensemble members.

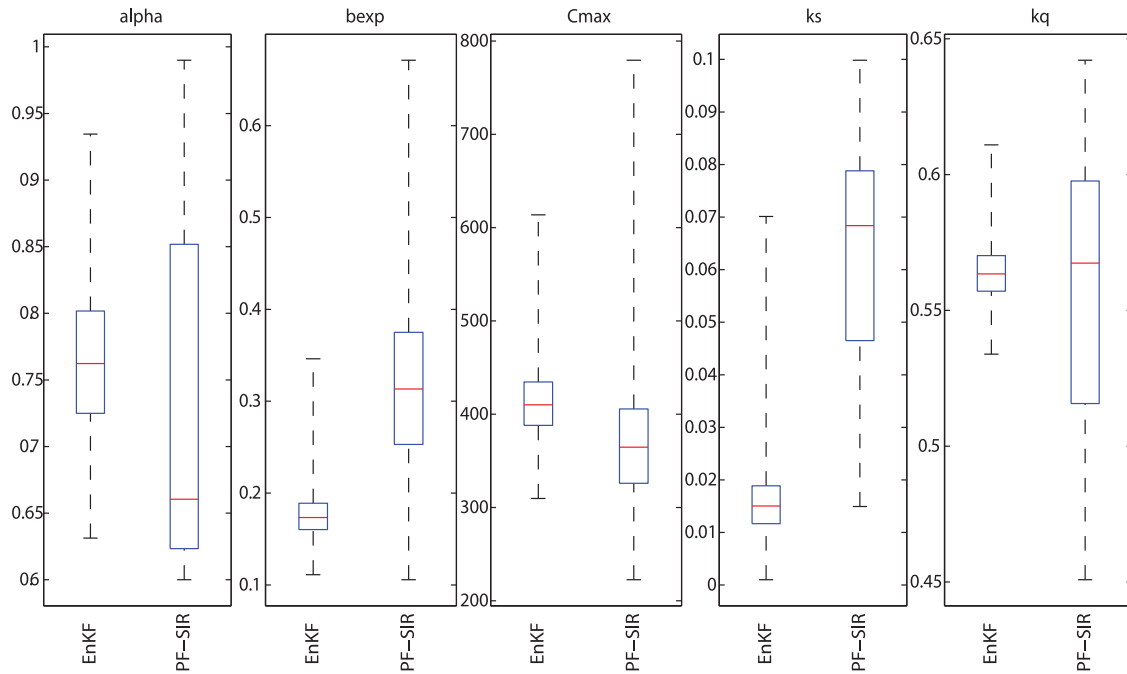


Figure 9. Box and whisker plots of the posterior distribution of each parameter using the EnKF and PF-SIR in the HyMod model.

model, the reader is directed to the validation results in Table 4. Interestingly, both techniques showed similar accuracy in the expected value of prediction. Though significant differences are found in the ability of each method to estimate the posterior distribution, the EnKF and PF-SIR perform similarly in terms of expected value. Similar to the HyMod validation, the PF-SIR produced both a more accurate ensemble prediction and more accurate 95% predictive bounds, according to Table 4. It is also important to note that the PF-SIR is underconfident during the validation according to the NRR, but the EnKF remains overconfident. This provides further evidence of the tendency of the EnKF to overfit the model parameters. While the EnKF overfit the parameters, the PF-SIR estimated a more accurate posterior, which in this case led to underconfidence in the validation. Underconfidence in a validation scenario would be expected because a wider range of flows was observed during the 10 calibration time periods than during the one validation time period. This behavior is also suggested by the rank histograms and Q-Q plots in Figure 11. Similar to the results found in the HyMod model, it appears that the predictive distribution from the EnKF is more overconfident during the validation than the calibration. The reverse trend is observed when applying the PF-SIR. Since

the PF-SIR more accurately characterized the uncertainty in the parameters during the calibration, it can still effectively estimate the uncertainty during the validation. Overall, the predictive distribution produced by the PF-SIR appears to give a more accurate representation of the uncertainty than the EnKF, but both methods displayed similar ability to track the observation with an expected value. The consistently more accurate estimation of uncertainty in the PF-SIR, over multiple replicates in two separate models, suggests that it is a more robust estimator of uncertainty than the EnKF.

4.3. Divergence in the Ensemble Kalman Filter

[25] In section 4.3, an analysis of the divergence observed in the EnKF during the December 1967 to May 1973 model calibration is presented. Filter divergence can refer to two scenarios: slow loss of sensitivity of the model to the observation due to poorly defined error terms [*Houtekamer and Mitchell, 1998*], and catastrophic filter divergence in which the Gaussian assumption is violated because of extreme nonlinearities in the model, leading to severe overadjustments in the updates [*Harlim and Majda, 2010*]. The latter was observed in this experiment. In order to understand the problems associated with parameter estimation in the EnKF during this calibration period, it is important to compare the streamflow hydrograph produced by the EnKF and the lower zone tension water maximum (LZTWM) parameter evolution. Figure 12 shows that during 300 d of the EnKF calibration time period, the peaks in the EnKF prediction far overestimate the observation. This overestimation can be upward of $1000 \text{ m}^3 \text{ s}^{-1}$. On the time steps when this overestimation is noticeable, it is observed that the LZTWM has sudden spikes in the lower 95th percentile of its distribution. During these events, the LZTWM of several ensemble members (58 in the largest event) are

Table 4. Performance Measures for the SAC-SMA Model During Calibration and Validation

Performance Measure	Calibration		Validation	
	EnKF	PF-SIR	EnKF	PF-SIR
NSE	0.78	0.85	0.82	0.83
α	0.76	0.82	0.76	0.84
NRR	1.13	1.09	1.14	0.78
ER95 (%)	24	10	18	6

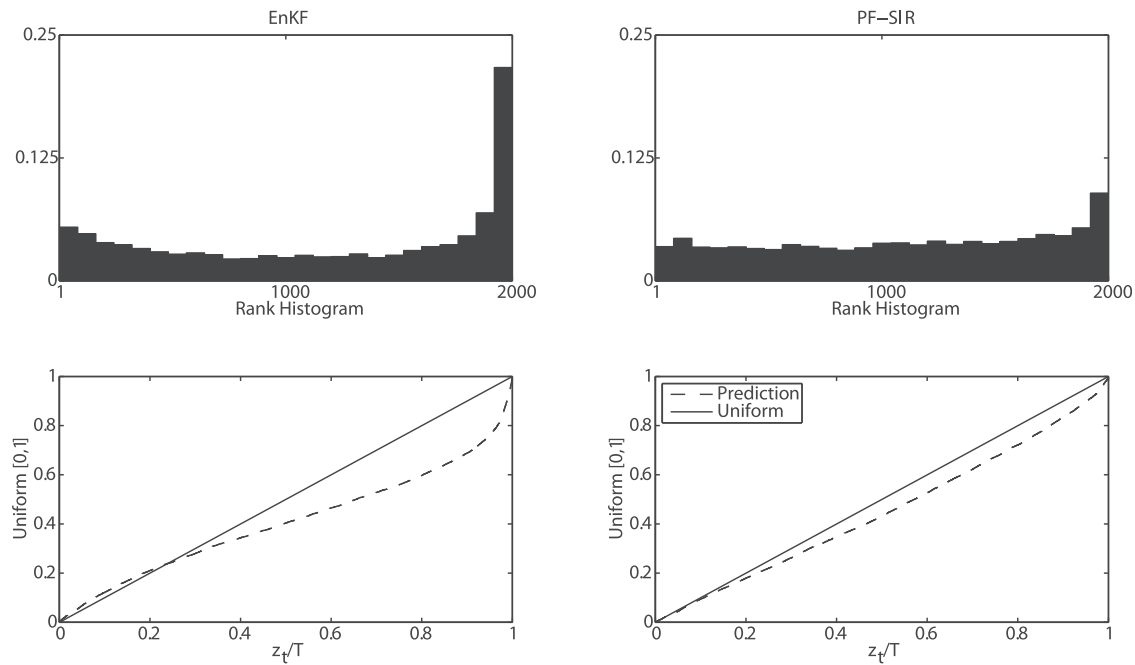


Figure 10. Rank histogram (top) and Q-Q plot (bottom) for the EnKF (left) and PF-SIR (right) for the SAC-SMA calibration runs with 2000 ensemble members.

adjusted from 500 mm to 10 mm (the maximum value to the minimum value). Since this parameter is a capacity of lower zone tension water storage, this sudden drop forced the given ensemble members to release excessive amounts of water, leading to significant overestimation of the streamflow. When examining the scenario for the PF-SIR, this phenomenon is not observed. Unlike the EnKF, which can

make large adjustments to state and parameter values in the event of large errors, the PF-SIR is more limited because it resamples, as opposed to adjusting, states and parameters. This makes the PF-SIR a more robust estimation technique, provided sufficient ensemble size. Though this is a rare occasion for the EnKF, as it was only observed once in this study and not well documented in the literature, it raises

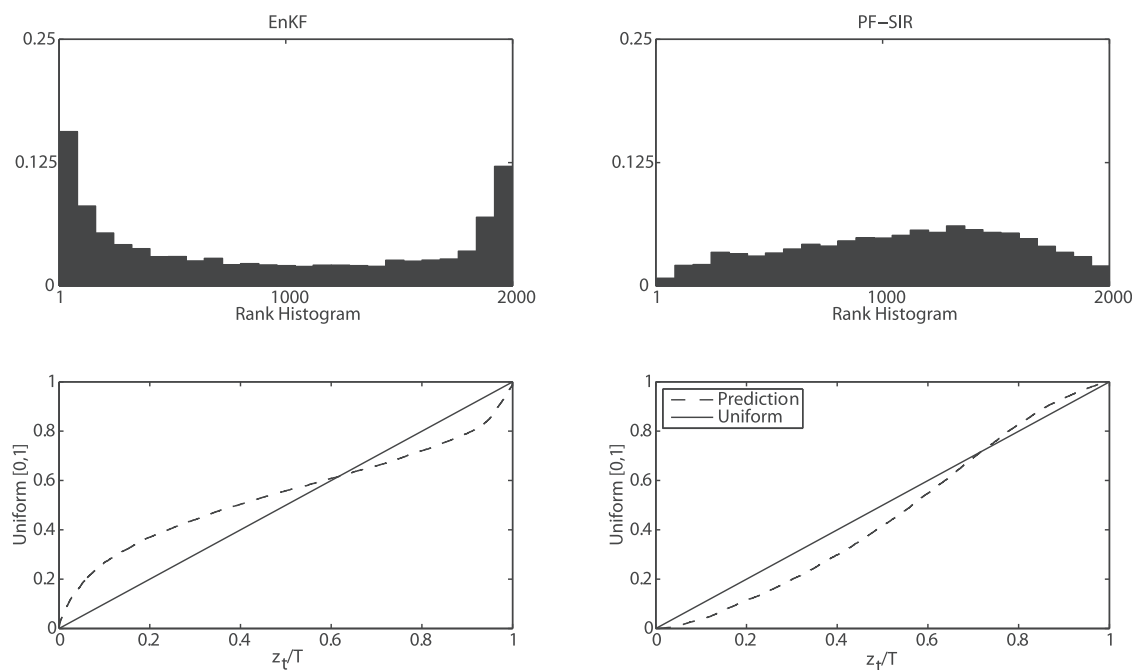


Figure 11. Rank histogram (top) and Q-Q plot (bottom) for the EnKF (left) and PF-SIR (right) for the SAC-SMA calibration runs with 2000 ensemble members.

questions about the confidence that can be placed on the EnKF to accurately estimate model parameters, in particular as the nonlinearity of the model increases.

4.4. Computational Time

[26] In addition to presenting results on the accuracy of the EnKF and PF-SIR with respect to ensemble size, this study presents results examining the computational requirements of the EnKF and PF-SIR, with respect to ensemble size, to illustrate the computational demands of each technique. The growth of computational demand with ensemble size for each technique is shown in Figure 13. From Figure 13, a trend not commonly presented in the literature is observed. This figure suggests that the EnKF, at each ensemble size, is more computationally demanding than the PF-SIR. In addition, larger ensemble sizes and an increased number of states and parameters, lead to a larger difference in computational demands between the EnKF and PF-SIR. This increased computational demand in the EnKF is caused by the calculation of covariances between predictions and all model states/parameters. While performing this calculation once is quite fast, over 2000 time steps and 1000 ensemble members this calculation can create a significant computational demand. The growth in computational demand for the PF-SIR is less steep because it is only necessary to calculate a weight for each ensemble member and perform resampling of the ensemble members. It is important to clarify that this figure is not presented to give the impression that the PF-SIR is more computation-

ally efficient than the EnKF, it is merely an illustration for the need to factor in the execution time of each technique, and not just the ensemble size, when determining which method is more efficient for a given application.

5. Discussion and Conclusion

5.1. Effects of Model Structure on Data Assimilation Techniques

[27] From the results obtained in this study, it is apparent that the model structures of the HyMod and SAC-SMA models have significantly different effects on the assimilation techniques. These differences are observed in both bias and uncertainty estimation. A bias in all of the results from the HyMod model was observed. This is different from the results in the SAC-SMA experiments, where a low bias was displayed for each method. In addition, the ability of each method to characterize the uncertainty in the model prediction of the HyMod and SAC-SMA are considerably different. It appears that with the increased complexity of the SAC-SMA model, uncertainty estimation, through data assimilation, is more accurate than appears to be possible in the HyMod model. This is somewhat intuitive as the increased number of parameters provides more flexibility in model structure. This comparison highlights that the accuracy of data assimilation techniques is model dependent and therefore model behavior must be examined when determining the effectiveness of a given data assimilation technique.

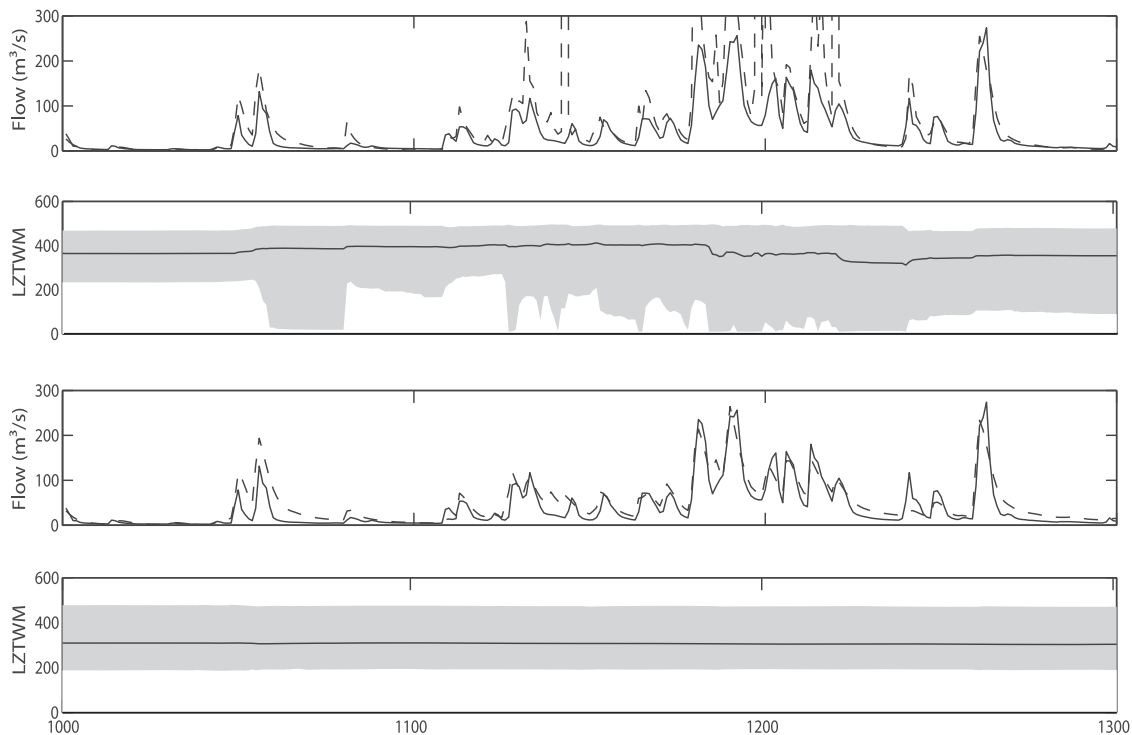


Figure 12. EnKF (top two) and PF-SIR (bottom two) calibrations starting in December 1967 of the prediction hydrograph and the LZW distribution. In the flow plots (first and third), the expected value is the dotted line and the observation is the solid line. For the parameter distribution plots, the dotted line is the 95% predictive bounds, the dashed line is the interquartile range, and the solid line is the expected value.

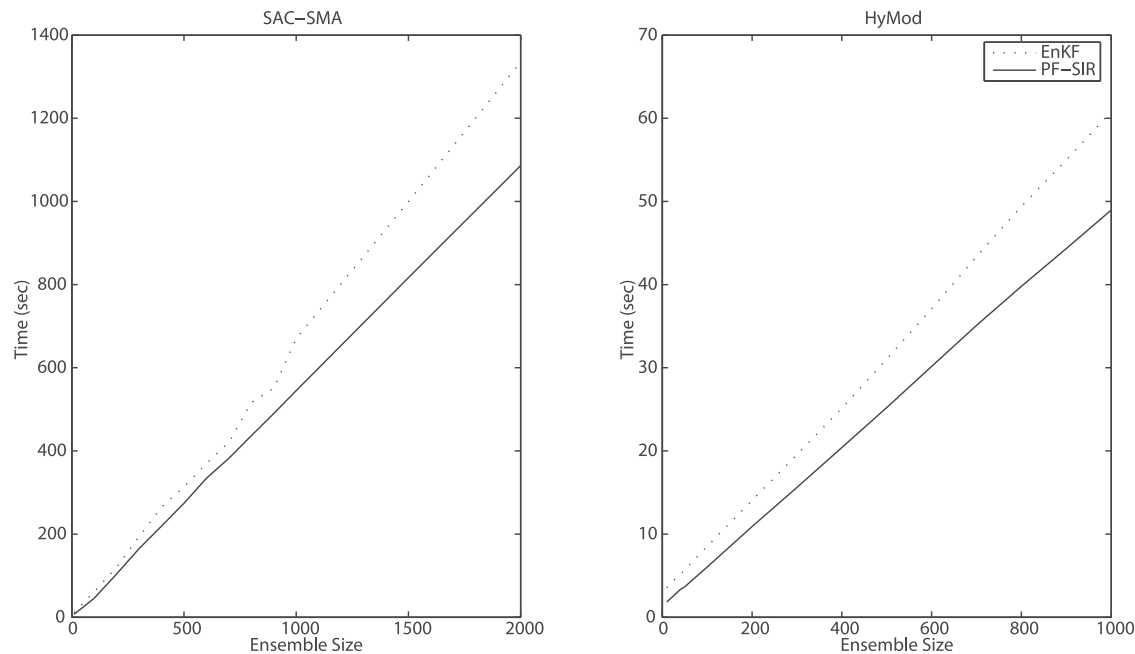


Figure 13. A comparison of the computational time as a function of ensemble size for the EnKF and PF-SIR.

5.2. Overconfidence and Divergence in the EnKF

[28] In the results presented, two problems of the EnKF were identified: a general trend toward overconfidence in the prediction of streamflow and a specific occasion of filter divergence. The cause of these errors in the EnKF can be inferred by comparison to the PF-SIR. While it is observed that both the EnKF and PF-SIR are overconfident in the HyMod model, results from the SAC-SMA model show that only the EnKF was overconfident. In addition, the EnKF was found to overfit the parameters during calibration in both models. This suggests a deficiency in the EnKF for prediction of both parameter and predictive uncertainty. Since the EnKF is poorly estimating the full posterior distribution, the error structure appears to be too skewed for reliable estimation of the full posterior distribution. Provided that the error structure is sufficiently non-Gaussian, the tails of the posterior distribution will be poorly estimated. This is found to be a consistent problem in the EnKF, but is less severe in the SAC-SMA model, where an error structure is likely to be less skewed than the HyMod model. Though the higher accuracy of the SAC-SMA model led to an increased ability of the EnKF to estimate uncertainty, the greater complexity led to difficulties in the linear estimation of states and parameters. Filter divergence was caused by a nonlinear relationship between the prediction and the LZTWM parameter under certain flow conditions. Since there is a sufficiently nonlinear relationship between this parameter and the prediction under these flow conditions, the Kalman update value was severely overestimated and several ensemble members were shifted to opposite ends of the parameter limits, leading to significant errors in streamflow estimation. Because this only occurred during one of the time-lagged replicates, the model is not sufficiently nonlinear to damage model predictions in most flow conditions and is therefore difficult to document.

Though it is rare, the potential for filter divergence raises questions about the robustness of the EnKF technique in increasingly nonlinear models.

5.3. Expected Value and Uncertainty

[29] In this study, verification of both techniques was performed through an analysis of the expected value and the predictive uncertainty to determine the benefits of each data assimilation method. It was important to analyze both expected value and predictive uncertainty to measure the ability of the model to track the observation, as well as represent the inherent uncertainty in the prediction. In section 4, contrasting results were obtained in comparing the accuracy of the expected value and uncertainty. In general, the EnKF and PF-SIR showed a similar ability to track the observation with the expected value, but differences were observed in uncertainty estimation. While the EnKF can be quite effective in predicting streamflow values, due to its restrictive assumptions, it struggles to predict uncertainty as accurately as the PF-SIR. This result highlights the importance of determining the goals of a study when implementing data assimilation on hydrologic models. A further conclusion is that if the goal is to track streamflow with an expected value, the EnKF may be able to perform this even at a smaller ensemble size, leading to higher computational efficiency, but the modeler must take precautions to ensure filter divergence does not occur. This result is consistent with previous studies [Zhou *et al.*, 2006; Nagarajan *et al.*, 2010; Weerts and El Serafy, 2006]. If quantification of the uncertainty in the prediction is important, the PF-SIR is likely a better choice. In general, it is suggested here that the characterization of uncertainty is important in most applications in hydrologic sciences and therefore needs to be discussed when using these techniques. The quantification of uncertainty is valuable from an operational and

research standpoint and should therefore be examined closely given the application.

[30] **Key Conclusions:** Both the EnKF and PF show similar abilities to track the observations; EnKF consistently produces overconfident results in comparison to the PF; and PF is a more robust parameter estimation technique than the EnKF.

[31] **Acknowledgments.** Partial financial support for this study was provided by NOAA-CPPA grant NA070AR4310203 and NOAA-MAPP grant NA110AR4310140. We would like to thank the three anonymous reviewers for their constructive comments that improved the clarity of this manuscript.

References

- Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249(1–4), 11–29, doi:10.1016/S0022-1694(01)00421-8.
- Bulygina, N., and H. Gupta (2009), Estimating the uncertain mathematical structure of a water balance model via Bayesian data assimilation, *Water Resour. Res.*, 45, W00B13, doi:10.1029/2007WR006749.
- Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system: Conceptual modeling for digital computers, *Tech. Rep.*, Joint Fed. and State River Forecast Center, U.S. Natl. Weather Serv. and Calif. Dept. of Water Resources, Sacramento, Calif., 204 pp.
- DeChant, C., and H. Moradkhani (2011a), Radiance data assimilation for operational snow and streamflow forecasting, *Adv. Water Resour.*, 34(3), 351–364, doi:10.1016/j.advwatres.2010.12.009.
- DeChant, C., and H. Moradkhani (2011b), Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, *Hydrol. Earth Syst. Sci.*, 15, 3399–3410, doi:10.5194/hess-15-3399-2011.
- Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, 53(4), 343–367, doi:10.1007/s10236-003-0036-9.
- Franssen, H. J. H., and W. Kinzelbach (2008), Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour. Res.*, 44, W09408, doi:10.1029/2007WR006505.
- Hamill, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, 129(3), 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.
- Harlim, J., and A. J. Majda (2010), Catastrophic filter divergence in filtering nonlinear dissipative systems, *Comm. Math. Sci.*, 8, 27–43.
- Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, 126, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Jeremiah, E., A. Sharma, S. Sisson, L. Marshall, and R. Mehrotra (2011), Bayesian calibration and uncertainty analysis of hydrological models: A comparison of adaptive-metropolis and sequential Monte Carlo samplers, *Water Resour. Res.*, 47, W07547, doi:10.1029/2010WR010217.
- Kavetski, D., S. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modelling, in *Calibration of Watershed Models*, *Water Sci. Appl. Ser.*, edited by Q. Y. Duan, et al., vol. 6, pp. 49–68, AGU, Washington, D. C.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrol. Earth Syst. Sci.*, 11(4), 1267–1277, doi:10.5194/hess-11-1267-2007.
- Leisenring, M., and H. Moradkhani (2011), Snow water equivalent estimation using Bayesian data assimilation methods, *Stoch. Environ. Res. Risk Assess.*, 25(2), 1–18, doi:10.1007/s00477-010-0445-5.
- Liu, Y., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.*, 43, W07401, doi:10.1029/2006WR005756.
- Montzka, C., H. Moradkhani, L. Weihermüller, H. Hendricks Franssen, M. Canty, and H. Vereecken (2010), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *J. Hydrol.*, 399(3–4), 410–421, doi:10.1016/j.jhydrol.2011.01.020.
- Moradkhani, H. (2008), Hydrologic remote sensing and land surface data assimilation, *Sensors*, 8, 2986–3004, doi:10.3390/s8052986.
- Moradkhani, H., and T. Meskele (2009), Probabilistic assessment of the satellite rainfall retrieval error translation to hydrologic response, in *Satellite Applications for Surface Hydrology*, edited by M. Gebremichael and F. Hossain, pp. 229–242, Springer, Water Sci. Technol. Library, N. Y., doi:10.1007/978-90-481-2915-7.
- Moradkhani, H., and S. Sorooshian (2008), General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis, in *Hydrological Modeling and Water Cycle: Coupling of the Atmospheric and Hydrological Models*, edited by S. Sorooshian et al., vol. 63, pp. 1–23, Springer, Water Sci. Technol. Library, N. Y., doi:10.1007/978-3-540-77843-1_1.
- Moradkhani, H., K. Hsu, H. Gupta, and S. Sorooshian (2005a), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, *Water Resour. Res.*, 41, W05012, doi:10.1029/2004WR003604.
- Moradkhani, H., S. Sorooshian, H. V. Gupta, and P. R. Houser (2005b), Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv. Water Resour.*, 28(2), 135–147, doi:10.1016/j.advwatres.2004.09.002.
- Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian (2006), Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, *Geophys. Res. Lett.*, 33, L12401, doi:10.1029/2006GL026855.
- Nagarajan, K., J. Judge, W. D. Graham, and A. Monsivais-Huertero (2010), Particle filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions, *Adv. Water Resour.*, 34, 433–447, doi:10.1016/j.advwatres.2010.09.019.
- Parrish, M., H. Moradkhani, and C. M. DeChant (2012), Towards reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resour. Res.*, 48, W03519, doi:10.1029/2011WR011116.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Rings, J., J. A. Huisman, and H. Vereecken (2010), Coupled hydrogeophysical parameter estimation using a sequential Bayesian approach, *Hydrol. Earth Syst. Sci.*, 14(3), 545–556, doi:10.5194/hess-14-545-2010.
- Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, 376(3–4), 428–442, doi:10.1016/j.jhydrol.2009.07.051.
- Smith, P. J., K. J. Beven, and J. A. Tawn (2008), Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, *Water Resour. Res.*, 44, W01410, doi:10.1029/2006WR005205.
- Smith, T. J., and L. A. Marshall (2008), Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques, *Water Resour. Res.*, 44, W00B05, doi:10.1029/2007WR006705.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Wang, D., Y. Chen, and X. Cai (2009), State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter, *Water Resour. Res.*, 45, W11416, doi:10.1029/2008WR007401.
- Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resour. Res.*, 42, W09403, doi:10.1029/2005WR004093.
- Wu, W., J. S. Clark, and J. M. Vose (2010), Assimilating multi-source uncertainties of a parsimonious conceptual hydrological model using hierarchical Bayesian modeling, *J. Hydrol.*, 394(3–4), 436–446, doi:10.1016/j.jhydrol.2010.09.017.
- Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the performance of the ensemble Kalman filter for land surface data assimilation, *Mon. Weather Rev.*, 134, 2128–2142, doi:10.1175/MWR3153.1.

C. M. DeChant and H. Moradkhani, Department of Civil and Environmental Engineering, Portland State University, 1930 SW 4th Ave., Ste. 200, Portland, Oregon, USA. (hamidm@cecs.pdx.edu)